

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-198680

(43)Date of publication of application : 31.07.1998

(51)Int.Cl.

G06F 17/28

G06F 17/30

(21)Application number : 09-000547

(71)Applicant : HITACHI LTD

(22)Date of filing : 07.01.1997

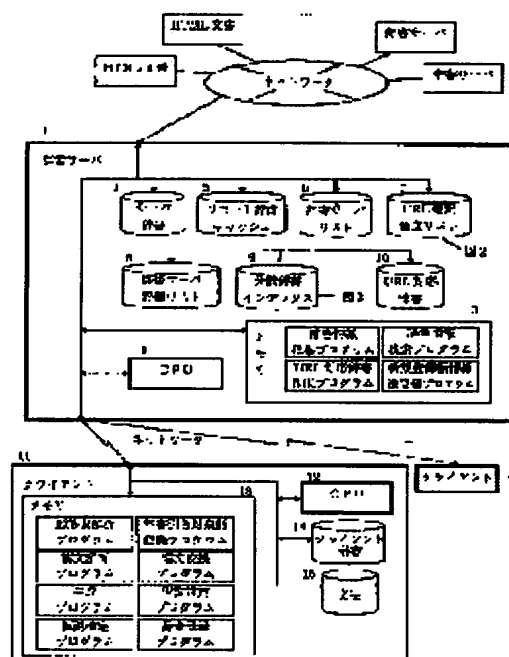
(72)Inventor : MORIMOTO YASUTSUGU  
MATSUDA JUNICHI

## (54) DISTRIBUTED DICTIONARY MANAGING METHOD AND MACHINE TRANSLATING METHOD USING THE METHOD

## (57)Abstract:

**PROBLEM TO BE SOLVED:** To provide a high-quality translated sentence by retrieving the translation knowledge of retrieval requested syntax unit, returning it to a client and preparing the translated sentences of whole object document at the time point when the translation knowledge is returned from a dictionary server.

**SOLUTION:** Plural dictionary servers 1 and clients 11 exist on a network. The client 11 can singly execute translating processing while utilizing a client dictionary 14 and grammar 15 and further, the retrieval of dictionary information concerning an index word not existent in the client dictionary 14 is requested to the dictionary server 1 at the time of translation execution. Then, the translation knowledge of retrieval requested syntax unit is retrieved and returned to the client 11 and at the time point when the translation knowledge is returned from the dictionary server 1, the client 11 prepares the translated sentences of whole translation object document while including translation disabled parts. In this case, the translation is executed while using the returned dictionary information and the client dictionary 14 together.



## LEGAL STATUS

[Date of request for examination] 05.04.2002

[Date of sending the examiner's decision of rejection] 26.08.2003

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision]

of rejection]

[Date of requesting appeal against examiner's  
decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-198680

(43) 公開日 平成10年(1998) 7月31日

(51) Int.Cl.<sup>6</sup>

G 0 6 F 17/28  
17/30

識別記号

F I

G 0 6 F 15/38  
15/40

C

3 1 0 C  
3 7 0 J

審査請求 未請求 請求項の数18 O L (全 20 頁)

(21) 出願番号 特願平9-547

(22) 出願日 平成9年(1997) 1月7日

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 森本 康嗣

東京都国分寺市東恋ヶ窪一丁目280番地

株式会社日立製作所中央研究所内

(72) 発明者 松田 純一

東京都国分寺市東恋ヶ窪一丁目280番地

株式会社日立製作所中央研究所内

(74) 代理人 弁理士 小川 勝男

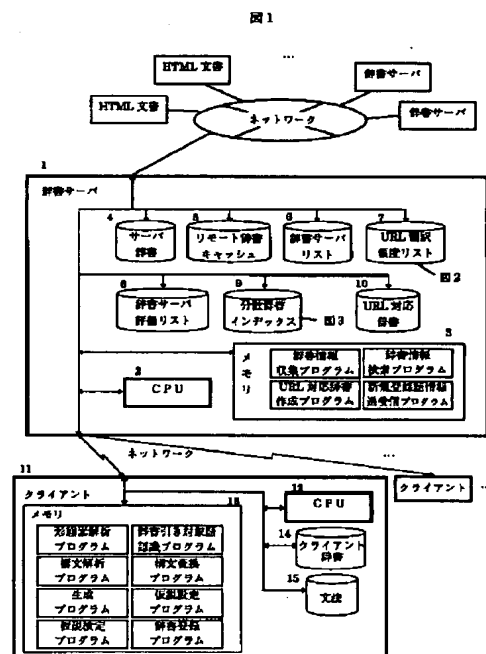
(54) 【発明の名称】 分散辞書管理方法及びそれを用いた機械翻訳方法

(57) 【要約】

【課題】 ネットワーク上に分散する辞書情報を共用することで質の高い訳語を得る。

【解決手段】 辞書サーバとクライアントがネットワーク上に複数存在し、各辞書サーバは、辞書情報の検索要求を受け、自身のサーバ辞書を検索し、サーバ辞書上に辞書情報がない場合には、他の辞書サーバに検索要求を行うことで、ネットワーク上に分散して存在する辞書情報をクライアントに返送する。

【効果】 過去の検索履歴およびネットワーク上の翻訳対象文書に対する翻訳要求の履歴を統計処理することによって、今後、検索される可能性が高い語や検索された見出し語を保持する可能性が高い辞書サーバを推定する。各クライアントは、クライアント上の辞書、文法では翻訳が不可能な場合に、ネットワーク上の辞書サーバを検索することにより、質の高い訳語を得る。その際、クライアント上での翻訳処理と辞書サーバの検索を非同期に行うことにより、ネットワークの速度が遅い場合でも、ユーザの使い勝手を損なうことなく辞書サーバの利用が可能になる。



## 【特許請求の範囲】

【請求項1】ネットワークに接続され、翻訳知識を保持する記憶手段を有する情報処理装置を用いて実行される機械翻訳方法であって、上記翻訳知識を参照して翻訳対象文書を翻訳する際に、上記翻訳対象文書中で上記翻訳知識では翻訳できない翻訳不可能部分を認識し、少なくとも該翻訳不可能部分を含み、文法的にまとまりのある構文単位に関する翻訳知識を該ネットワークに接続された他の情報処理装置に問い合わせ、上記構文単位に続く内容の翻訳を続行し、上記他の情報処理装置から上記翻訳不可能部分に関する翻訳知識を得て、上記翻訳不可能部分を含む上記翻訳対象文書の翻訳結果を作成することを特徴とする機械翻訳方法。

【請求項2】上記情報処理装置は上記翻訳不可能部分を含む構文単位について仮の翻訳処理を行った結果を出力し、上記他の情報処理装置から得られた翻訳知識に基づいて翻訳結果を作成する処理では、該仮翻訳結果を上記翻訳知識に基づいて変更することを特徴とする請求項1の機械翻訳方法。

【請求項3】上記仮の翻訳処理では、上記翻訳対象文書における未知語の品詞を推定し、該推定された品詞に従って翻訳処理を実行し、翻訳結果を変更する処理では、該推定された品詞と他の情報処理装置から得られた該未知語の品詞を比較し、一致していれば訳語のみを変更し、一致していなければ該未知語を含む文全体を際翻訳することを特徴とする請求項2に記載の機械翻訳方法。

【請求項4】機械翻訳装置における翻訳結果の表示方法であって、翻訳対象文書を翻訳する際に、その時点で利用可能な翻訳知識を用いて翻訳を実行して、翻訳結果を出力し、利用可能な翻訳知識に変更があったかどうかを認識し、変更があった場合には、該変更に基づいて過去の翻訳結果の出力を変更することを特徴とする機械翻訳結果の表示方法。

【請求項5】ネットワークに接続され、辞書と文法を保持する記憶手段を有する情報処理装置を用いて実行される機械翻訳方法であって、上記辞書と文法を参照して翻訳対象文書を翻訳する際に、上記翻訳対象文書中で上記辞書に存在しない未知語の有無を認識し、該未知語が存在しない場合には上記辞書のみを用いて翻訳を実行し、該未知語が存在した場合にはネットワークに接続された他の情報処理装置が有する辞書への検索要求を行い、その結果得られた情報を用いて上記翻訳対象文書の翻訳を完成させるとを特徴とする機械翻訳方法。

【請求項6】ある見出し語に関する辞書検索要求を受け取り、該見出し語に関する辞書情報をサーバ辞書を検索して取得し、該見出し語に関する辞書情報が前記サーバ辞書に存在しない場合には、他辞書サーバに検索要求を行い、他辞書サーバから取得した辞書情報を返送する分散辞書検索方法において、過去の検索要求に関する情報に基づいて他辞書サーバの優先順位を決定し、該優先順

位に従って他辞書サーバに検索要求を行うことを特徴とする分散辞書検索方法。

【請求項7】ある見出し語に関する辞書検索要求を受け取り、該見出し語に関する辞書情報をサーバ辞書を検索して取得し、該見出し語に関する辞書情報が前記サーバ辞書に存在しない場合には、他辞書サーバに検索要求を行い、他辞書サーバから取得した辞書情報を返送する分散辞書検索方法において、翻訳対象文書中に出現する単語に基づいて他辞書サーバの優先順位を決定し、該優先順位に従って他辞書サーバに検索要求を行うことを特徴とする分散辞書検索方法。

【請求項8】上記他辞書サーバの優先順位決定方法は、上記翻訳対象文書中に出現する単語の分野情報に基づいて上記翻訳対象文書の分野を推定し、該推定された分野における過去の検索要求に関する情報に基づいて他辞書サーバの優先順位を決定することを特徴とする請求項7に記載の分散辞書検索方法。

【請求項9】上記他辞書サーバの優先順位決定方法は、上記翻訳対象文書中に出現する少なくとも一部の単語の辞書情報を保持する他辞書サーバの情報を保持し、該他辞書サーバの情報および上記翻訳対象文書中の単語を用いて他辞書サーバの優先順位を決定することを特徴とする請求項7に記載の分散辞書検索方法。

【請求項10】ネットワーク上に存在する文書に対する翻訳頻度情報を利用して、ネットワーク上に存在する文書の重要度を決定し、該重要なネットワーク文書を検索し、該文書中の未知語を認識し、該未知語を辞書登録対象語の候補として出力することを特徴とする辞書構築支援方法。

【請求項11】ネットワーク上に存在する文書に対する翻訳頻度情報を利用して、ネットワーク上に存在する文書の重要度を決定し、該重要なネットワーク文書を検索し、該文書中の未知語を認識し、該未知語に関する辞書情報をネットワーク上の他辞書サーバに対して要求し、得られた辞書情報を辞書ファイルに格納する辞書構築方法。

【請求項12】ネットワークに接続され、少なくとも辞書情報を保持する記憶装置を有する情報処理装置において、辞書情報の新規登録を検出し、該新規登録された語の見出しの情報と自情報処理装置の識別子のみを他の情報処理装置に送信し、該見出しの情報を受信した情報処理装置では少なくとも該見出しの情報と識別子を格納することを特徴とする辞書情報配信方法。

【請求項13】ネットワークに接続され、少なくとも辞書を保持する記憶手段を有する情報処理装置において、上記辞書を参照して翻訳対象文書を翻訳する際に、上記翻訳対象文書中で上記辞書では翻訳できない未知語を認識し、該未知語に関する辞書情報を該ネットワークに接続された他の情報処理装置に問い合わせ、上記他の情報処理装置から上記未知語に関する辞書情報を得て、上記

未知語を含む上記翻訳対象文書の翻訳結果を作成し、上記辞書情報を上記辞書に格納することを特徴とする辞書構築方法。

【請求項14】見出し語と対応する少なくとも一つの訳語を保持する翻訳用辞書を利用する機械翻訳方法であって、翻訳対象文書を識別するための識別情報を保持し、該識別情報を用いて同一文書を翻訳する場合に見出し語に対する訳語が同一となることを特徴とする機械翻訳方法。

【請求項15】ある翻訳対象文書の機械翻訳結果における訳語の候補を表示し、ユーザによる訳語の変更情報を取得し、該変更情報に従って訳語を変更するとともに少なくとも該変更情報と上記翻訳対象文書の識別情報を記憶することを特徴とする機械翻訳用辞書の更新方法。

【請求項16】ネットワークに接続された文書を他の自然言語に翻訳するネットワーク分散型機械翻訳システムにおいて、

複数の辞書サーバを有し、各々の辞書サーバは、辞書とネットワークを介してアクセス可能な文書の統計情報に基づいて作成された分散辞書インデックスを有し、上記文書が含む語句の見出し語が上記辞書に存在しない場合には、上記分散辞書インデックスを参照して他の辞書サーバより該見出し語に関する情報を得て、該文書の翻訳を実行することを特徴とするネットワーク分散型機械翻訳システム。

【請求項17】ネットワークに接続された文書を他の自然言語に翻訳するネットワーク分散型機械翻訳システムにおいて、

複数の辞書サーバを有し、各々の辞書サーバは、辞書とネットワークを介してアクセス可能な文書のアクセス頻度情報に基づいて作成された分散辞書インデックスを有し、上記文書が含む語句の見出し語が上記辞書に存在しない場合には、上記分散辞書インデックスを参照して他の辞書サーバより該見出し語に関する情報を得て、該文書の翻訳を実行することを特徴とするネットワーク分散型機械翻訳システム。

【請求項18】辞書を参照しながら自然言語によって表される文書を他の種類の自然言語に翻訳するコンピュータシステムで用いられるコンピュータ可読媒体であって、

自然言語で表された翻訳対象文書を読み込むための第1の命令コードと、

上記翻訳対象文書から辞書引き対象語句を抽出するための第2の命令コードと、

上記辞書引き対象語句の意味を予め準備された辞書から得て他の自然言語に翻訳するための第3の命令コードと、

上記辞書が上記辞書引き対象語句の意味を保持しないときに、該辞書引き対象語句を未知語と認定して該未知語部分の翻訳を停止したまま後続部分の翻訳を継続させる

ための第4の命令コードと、

ネットワークで接続された他の情報端末に対して上記未知語に関する情報を問い合わせるための第5の命令コードと、

上記他の情報端末から得られた上記未知語に関する情報によって、上記未知語部分の翻訳を確定し、他の翻訳部分と併せて翻訳結果とするための第6の命令コードとを有するコンピュータプログラムを保持することを特徴とするコンピュータ可読媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、ネットワーク上に分散して存在する対訳辞書を管理する方法およびそれを利用した機械翻訳方法に関する。

【0002】

【従来の技術】現在、ネットワークの発達、特にインターネットの普及に伴って、主に英語を中心とする外国語の情報に容易にアクセスできるようになっている。そのため、例えば、日経ホーム出版社、日経トレンディ、1995年9月号、pp.111-121に見られるようにWWW(World Wide Web)を介して得られるHTML(HyperText Markup Language)文書を翻訳するシステムが開発されている。

【0003】また、ネットワークを利用して機械翻訳の辞書を共同で利用する例として、特開平6-259600号がある。これは、ネットワークを介して辞書を分散した辞書分散システムを提案する。翻訳に適用する場合は、英語・露語といった各国言語の翻訳辞書を分散させておき、入力された文字が何語であるかを識別した後に、該当する翻訳辞書を有するセンタに入力データ(文字のイメージデータ)を転送する。このような従来技術では、特定の分野や特定の言語を翻訳する辞書は特定されており、その辞書で翻訳できない語は翻訳不可能文字として処理される。

【0004】

【発明が解決しようとする課題】機械翻訳において、辞書の質とカバー率は、翻訳の質に大きく影響する。しかし、インターネット上の情報は、非常に幅広い分野をカバーしており、また、日々新しい情報が現われるため、広範な分野の語句や新語などに対応して辞書を増補していくことに非常にコストがかかるという問題点があった。

【0005】また、辞書を提供する立場の機械翻訳メーカーとしては、従来のように数ヶ月おきに新しいバージョンの辞書を提供するといったやり方では、ユーザに必要な辞書情報をタイムリーに提供することができないという問題点があった。

【0006】翻訳処理では、辞書へのアクセスが頻繁に行われるため、辞書アクセスの速度は、処理速度に大きく影響を与える。インターネットでは、回線速度が速いことが期待できないため、翻訳処理中に、単にインター

ネットを介して、多くの辞書にアクセスすると翻訳が高速に完了することはない。また、現在、一般家庭からインターネットに接続するためには、電話回線のようにLANなどと比較すると非常に低速な回線を利用するため、ネットワークを介した辞書アクセスは非常に時間がかかり一般家庭向きではない。

【0007】本発明の第1の目的は、クライアントが、翻訳処理を行う際に、ネットワークを介した辞書検索の処理速度が十分ではない場合でも、ネットワークを介して辞書情報を検索して高品質な訳文を作成できる機械翻訳方法を提供することにある。

【0008】本発明の第2の目的は、辞書サーバにおいて、クライアントから辞書情報検索要求がある前に、次に検索要求が来る可能性が高い未知語を予測することによって、辞書構築作業を支援し、辞書情報検索要求が来た時点でレスポンスを高速にすることができる辞書サーバの辞書情報収集方法を提供することにある。

【0009】本発明の第3の目的は、辞書サーバにおいて、辞書情報検索要求があった時点で、検索要求のあった見出し語が辞書サーバ上にない場合、その見出し語の辞書情報を持つ、または、持つ可能性が高い他の辞書サーバに高い優先順位を与え、優先順位の高い順に検索要求を行うことで、所望の辞書情報を検索するためのコストを抑えることができる辞書サーバを提供することにある。

【0010】本発明の第4の目的は、ユーザが必要とする辞書情報のみをタイムリーに配信することができる辞書情報配信方法を提供することにある。

【0011】本発明の第5の目的は、一般的な情報処理装置に適用して自然言語を他の自然言語に機械翻訳処理を行わせる処理手順（プログラム、命令コードの集合体）や必要なデータを格納したコンピュータシステムで読みとり可能な記録媒体を提供することにある。

【0012】  
【課題を解決するための手段】上記第1の目的を達成するため本発明は、ネットワークに接続されたクライアントおよび辞書サーバからなる翻訳システムにおいて、クライアントにおいて翻訳処理を実行中に未知語などの翻訳不可能部分を認識し、上記翻訳不可能部分を含む構文的にまとまりのある構文単位についての翻訳知識の検索を辞書サーバに要求し、上記構文単位に続く内容の翻訳を続行し、一方、辞書サーバでは、検索要求のあった構文単位についての翻訳知識を検索してクライアントに返送し、辞書サーバから翻訳知識が返送された時点で、クライアントでは、上記翻訳不可能部分を含む翻訳対象文書全体の訳文を作成する。

【0013】上記第2の目的を達成するため本発明は、ネットワークに接続された辞書サーバにおいて、ネットワーク上に存在する文書への翻訳頻度情報を保持し、上記翻訳頻度情報を用いて上記文書の重要度を決定し、上

記重要度の高い文書を検索し、該重要度の高い文書中の未知語を認識・出力する。

【0014】上記第3の目的を達成するため本発明は、ネットワークに接続された辞書サーバにおいて、ネットワーク上に存在する他の辞書サーバへの辞書情報の検索要求に対する結果の履歴を保持し、クライアントからある見出し語に関する辞書情報検索要求が行われた際、自辞書サーバ上の辞書に上記見出し語の辞書情報が存在しない場合には、上記履歴を用いて上記見出し語を保持する／保持する可能性が高い辞書サーバに高い優先順位を与え、優先順位の高い順に他辞書サーバに上記見出し語に関する辞書情報の検索要求を行う。

【0015】上記第4の目的を達成するため本発明は、ネットワークに接続されたクライアントおよび辞書サーバからなる翻訳システムにおいて、クライアントにおいて翻訳処理を実行中に未知語などの翻訳不可能部分を認識し、上記翻訳不可能部分を含む構文的にまとまりのある構文単位についての翻訳知識の検索を辞書サーバに要求し、一方、辞書サーバでは、検索要求のあった構文単位についての翻訳知識を検索してクライアントに返送し、辞書サーバから翻訳知識が返送された時点で、該翻訳知識をクライアント上の辞書に格納する。

【0016】上記第5の目的を達成するために本発明のコンピュータ可読媒体は、翻訳対象文書を読み込むプログラムと、翻訳対象文書の中から辞書引き対象語句を抽出するプログラムと、辞書引き対象語句の意味を調べて他言語に変換するプログラムと、ローカルに参照する辞書中、つまり翻訳時に第1にアクセスする辞書中に調べたい語句の見出しが存在しない場合には、その意味の不明な語句部分の翻訳をスキップして後続の部分の翻訳処理を続行させるプログラムと、ネットワークを介して不明な語句部分の意味を他の情報処理端末に問い合わせるプログラムと、不明な語句の情報が得られた時にその不明語句部分の翻訳を確定させて他の翻訳済みの部分と併せて最終結果として出力するプログラム、を格納するように構成されている。

【0017】（実施例）本発明の実施例を図面を用いて説明する。

【0018】（1）システム全体構成

図1に、本実施例の全体の構成を示す。辞書サーバ1およびクライアント11がネットワーク上に複数存在する。クライアント11は、クライアント辞書14および文法15を利用して単独で翻訳処理を実行可能であり、さらに翻訳実行時に、クライアント辞書14に存在しなかった見出し語についての辞書情報の検索を辞書サーバ1に対して要求し、返送された辞書情報とクライアント辞書14を併せて用いて翻訳を実行する。辞書サーバ1は、ある見出し語に対する辞書情報検索要求をクライアント11から受け取り、その見出し語に関する辞書情報を返送する。

【0019】以下、辞書サーバ1の構成について説明す

る。CPU2は、メモリ3および各種ファイルにアクセスし、処理を行う。メモリ3には、辞書サーバ1上で各種の処理を行うために用いるプログラムが格納される。

【0020】サーバ辞書4は、辞書サーバ1に備えられた翻訳用辞書であり、クライアントからの辞書情報検索要求があった場合には、まずこの辞書が検索される。ここでは、サーバ辞書4が1個しかない例で説明しているが、実際には、複数のファイルに分かれていても構わない。

【0021】リモート辞書キャッシュ5は、ネットワーク上に存在する他の辞書サーバからの検索結果を格納しておく辞書ファイルである。リモート辞書キャッシュ5を用いることにより、再度、同じ見出し語について検索要求があった場合に、他の辞書サーバに対して検索要求を行うことなく、辞書情報を取得できる。

【0022】辞書サーバリスト6は、ネットワーク上にある他の辞書サーバのリストである。ただし、ネットワーク上の全ての辞書サーバが記載されている訳ではない。

【0023】URL翻訳頻度リスト7は、翻訳依頼のあったHTML文書のURLと各URLに対する単位時間当たりの翻訳要求の頻度を表す。URL翻訳頻度リストの例を図2に示す。図2の場合、1日当たりの翻訳要求の頻度を示している。つまり、例えば、1のURL「http://info.../robots.html」については、X月Y日に60回の翻訳依頼があったことを示す。

【0024】辞書サーバ評価リスト8は、辞書サーバリスト6中の各辞書サーバ毎に、過去の検索要求に関する履歴情報が格納されている。

【0025】分散辞書インデックス9は、サーバ辞書4、リモート辞書キャッシュ5、他の辞書サーバを含めて、辞書の見出し語情報を管理するインデックスファイルである。分散辞書インデックス9の構造を図3に示す。分散辞書インデックス9は、見出し語と各見出し語に対する辞書情報が格納されている箇所を示す辞書情報ポインタからなる。辞書情報ポインタは、辞書情報ポインタの種別とポインタの値からなる。辞書情報ポインタの種別は、辞書情報の格納先を示しており、サーバ辞書の場合はL、リモート辞書キャッシュの場合はC、それ以外の場合はRである。ポインタ値の意味は、種別によって異なる。種別がLの場合、辞書情報ポインタは、ローカル辞書中の辞書データのバイト位置を示している。種別がCの場合、辞書情報ポインタは、リモート辞書キャッシュ中の辞書データのバイト位置を示している。種別がRの場合、辞書情報ポインタは、辞書サーバID（ここでは、IPアドレス）を示している。例えば、見出し語Big Blueの場合、辞書情報は2箇所格納されている。1つは、リモート辞書キャッシュの先頭から44825バイトの位置から開始するデータとして、もう1つは、IPアドレスが133.144.18.152の辞書サーバに格納されている。

【0026】URL対応辞書10は、URL毎に個別の辞書情報が登録された辞書である。例えば、president→「大統領」、「社長」のように、複数の訳語を持つ見出し語の場合、あるURLにおいてpresidentが「社長」の意味でしか用いられない場合には、president→「社長」という辞書情報をこのURL対応辞書10に格納しておく。

【0027】クライアント11においても、CPU12がメモリ13とファイルにアクセスして処理を実行するのは辞書サーバ1と同様である。メモリ13上には、クライアント側で翻訳を実行するプログラム及びデータが格納されている。クライアント辞書14および文法15は、翻訳処理に利用されるデータである。また、辞書サーバから検索された辞書情報はクライアント辞書14に格納される。これにより、その後に同じ単語が出現した場合には、辞書サーバへの検索要求が不要になり、また、分野のような固定的な分類に囚われることなく、ユーザが必要とする単語の情報を入手することが可能となる。

【0028】(2) 辞書サーバにおける未知語収集次に、辞書サーバ1における辞書情報管理方法について説明する。各辞書サーバは、辞書情報管理のために、ネットワークを介して他の辞書サーバと情報のやりとりを行う。本実施例では、次の2種類の方法で辞書サーバ間の情報のやりとりを行う。

【0029】まず、第1の方法として、辞書サーバ1における新規登録語の送受信処理について説明する。

【0030】各辞書サーバは、サーバ辞書4に新規に登録された語に関する情報を他の辞書サーバに送信する。送信相手となる辞書サーバは、辞書サーバリスト6を参照して決定する。新規登録語情報を受信した辞書サーバでは、新規登録語を保持している辞書サーバのIDを知ることができる。

【0031】以下、図4を用いて、辞書サーバにおける新規登録語情報送信処理について説明する。

【0032】辞書サーバ1において、サーバ辞書4に新規見出し語の辞書情報の登録が発生すると、新規見出し語を新規見出し語リスト(図示せず)に登録する(ステップ11)。新規登録は、辞書サーバの管理者が行っても良いし、ユーザが作成した辞書を辞書サーバに登録しても良い。メモリ3上にロードされた新規登録語情報送受信プログラムは、タイマーによって一定時間おきに起動され(ステップ12)、新規見出し語リスト中に見出し語が登録されているかどうかをチェックする(ステップ13)。登録されていれば、新規登録語情報送受信プログラムは、辞書サーバリスト6中の辞書サーバに新規登録語情報を送信する(ステップ14)。新規登録語情報は、図5のような構造を持つ。図5において、辞書サーバIDは、新規登録語情報を送信した辞書サーバのID番号(IPアドレス)であり、さらに新規登録された語の見出し語のリストが並ぶ。

【0033】他の辞書サーバでは、図6に示す手順によ

って、新規登録語受信処理を行い、見出し語を登録する。他の辞書サーバは、送信された登録見出し語情報を受信したかどうか判別し(ステップ15)、受信した場合には、見出し語登録処理を行う(ステップ16)。即ち、受信した新規登録語情報と辞書を比較し、未知語であった場合には、新規登録語情報中の見出し語情報とサーバのID番号を分散辞書インデックス9に格納する。このようにして、各辞書サーバは、他の辞書サーバが持っている辞書の見出し語情報を知ることができる。このとき、新規登録語情報を受信した辞書サーバにとっては、新規に登録される情報が見出し語とIPアドレスのみであるので、訳語も含めて登録する場合と比べて、辞書容量を節約できる。また、新規登録語情報を送信した辞書サーバにとっては、自分が新規に登録した辞書の訳語の情報を無断で複写されたりすることを防止できる。これにより、辞書サーバ間で課金を行う場合にも本方法を適用することが可能となる。

【0034】次に、第2の方法として、辞書サーバにおける辞書情報収集処理について説明する。辞書情報収集処理は、辞書サーバが未知語の辞書情報を自動的に収集する処理である。辞書情報収集処理では、今後利用される可能性が高い語を統計情報を用いて自動的に推定し、推定された語の辞書情報を持つ可能性が高い他の辞書サーバを統計情報を用いて自動的に推定することにより、通信コストを抑えつつ、必要な辞書情報を収集することに特徴がある。

【0035】以下、図7を用いて、辞書情報収集処理の全体の処理フローを説明する。

【0036】タイマーによって時間を監視し、辞書情報収集プログラムを一定時間おきに起動する(ステップ21)。辞書情報収集プログラムは、URL翻訳頻度リスト7に基づいて今後翻訳される頻度が高い高翻訳頻度文書を予測する(ステップ22)。この予測は、文書のURLに対する辞書検索要求の回数を統計的に分析した結果に基づいて行う。例えば、本実施例では、以下の2通りの予測方法を説明する。

【0037】①単位時間当たりの辞書検索要求回数の平均が閾値以上であるURL

②単位時間当たりの辞書検索要求回数の変化率が閾値以上であるURL

ここで、①は、例えば、継続的に人気のある雑誌のホームページのように、常に多くの翻訳要求がある文書を抽出するための方法である。②は、新設されたURLで人気が高いものを抽出するための方法である。

【0038】ここでは閾値以上のURLを抽出するように設定したが、他にも、翻訳頻度の平均および変化率が高い順に予め定められた個数のURLを対象とするなどの方法をとっても良い。また、URL翻訳頻度リストには、実際の翻訳頻度以外に、最優先であることを示すコード(例えば、-1)を人手により設定しておき、実際の翻訳頻

度とは無関係に、常に未知語収集処理の対象とすることもできる。

【0039】なお、本実施例では、URLに対する翻訳要求の頻度に基づいて今後翻訳される可能性が高い文書の予測を行ったが、辞書サーバをWWWのproxy(代理)サーバと兼用させることができ、この場合には、URLへのアクセスの頻度情報を取得できるため、アクセス頻度情報とHTML文書の言語を示す情報から同様の処理を行うこともできる。

【0040】次に、翻訳頻度の高いURLを一つ選択し、該URLのHTML文書を取得する(ステップ23)。ここでは、全ての文書を取得する例を説明するが、他にも、前回アクセス時のHTML文書を保存しておき、差分を取って、処理対象とする方法や、予め定められた文字列(例えば、Whats New!など)で示されるHTML文書を取得して処理対象とする方法を採用することもできる。

【0041】HTML文書を取得したら、その文書中の辞書引き対象語を決定する(ステップ24)。ここで、辞書引き対象語となるのは、取得したHTML文書中に出現する単語のうちサーバ辞書4から正確な辞書情報が得られない単語である。ここでは、次の3つの場合を説明する。

【0042】①単語が辞書に登録されていない場合：機械翻訳においていわゆる未知語と呼ばれるものである。

【0043】②固有名詞の場合：単語としては、辞書に登録されているのだが、特別な意味を持つ場合がある単語の場合。

【0044】③複合語の場合：個々の単語については、辞書に登録されているのだが、複合語が特別な意味を持つ場合。例えば、Big Apple(ニューヨーク)のような複合語について、bigとappleは含まれているが、Big Appleという複合語が含まれていない場合。

【0045】辞書引き対象語抽出処理については、後で詳細に説明する(図10)。

【0046】辞書引き対象語が決定すると、他の辞書サーバの優先度を決定する(ステップ25)。なお、本ステップ25については、後で詳細に説明する(図10)。

【0047】他の辞書サーバに対して、決定した優先度に従い辞書引き対象語に関する辞書情報の検索要求を行う(ステップ26)。ステップ26で用いる検索要求データの構造を図8に示す。検索要求データは、検索対象種別、検索深さ、見出し語、URL、検索時刻からなる。検索対象種別は、LかRの2種類である。Lの場合は、検索要求を受けた辞書サーバ上の辞書、即ち、サーバ辞書とリモート辞書キャッシュのみを検索対象とし、Rの場合は検索要求を受けた辞書サーバ上の辞書と他辞書サーバの両方を検索対象とする。検索深さは、あるクライアントからの辞書検索要求に対し、いくつの辞書サーバを経由して、検索要求が来ているかを示している。ある辞書サーバが平均して、n個の辞書サーバに検索要求を出すとなると、検索要求を受けるサーバの数は、検索深さの指数



関数のオーダとなるため、トラフィックが急激に増大する。そのため、予め定めた値より大きい深さの検索要求は行わないようにする。見出し語は、検索対象となる見出し語の文字列である。URLは、翻訳対象文書がネットワーク上のHTML文書である場合には、その文書のURLが格納されており、それ以外の場合には、デフォルト値が設定される。検索時刻は、該見出し語について、クライアントからの最初の検索が行われた時間が格納されており、最初の検索から一定の時間が経過した検索要求は破棄される。あるいは、クライアントにおいて、検索が行われた時刻ではなく、検索結果を返送する締め切りの時刻を設定し、その時刻に間に合わなかった検索要求を破棄するようにしても良い。

【0048】ステップ26における検索結果返送データを受信し、検索が成功した場合には、辞書データを更新する(ステップ27)。検索結果返送データの構造を図9に示す。検索結果返送データは、辞書サーバ名、成功フラグ、内容種別、暗号化フラグ、辞書情報からなる。辞書サーバ名はこの辞書情報を格納している辞書サーバIDである。成功フラグは、検索が成功したかどうかを示すフラグである。内容種別は、辞書情報の内容の種別を表し、Eの場合は、該見出し語が辞書サーバIDが示す辞書サーバ中に存在したことを示し、Cの場合には、該見出し語に関する辞書情報そのものが得られたことを示す。暗号化フラグは、この辞書情報が暗号化されているかどうかを示すフラグである。

【0049】ここで、検索が成功した場合に、次の3通りの処理を行う。第1に、他辞書サーバが暗号化されていない辞書情報そのものを送付した場合には、辞書情報を複写したものをリモート辞書キャッシュに格納し、分散辞書インデックス9を更新する。第2に、その見出し語に関する辞書情報の有無のみが送付された場合には、分散辞書インデックス9に該見出し語を追加し、辞書サーバの名称をポインタとして格納する。

【0050】第3に、暗号化された辞書情報が得られた場合には、リモート辞書キャッシュに格納する方法とデータを破棄する方法がある。クライアントのユーザが暗号化された辞書情報を復号する鍵を持っていることが分かっている場合、例えば、辞書サーバ同士で辞書情報を交換することが予め取り決められている場合には、暗号化されたデータをそのまま保持しておき、クライアントからの検索要求に従って、辞書サーバIDとともに辞書情報を送付することでクライアント側で復号した辞書情報を利用できる。それ以外の場合には、辞書情報を利用できないので、データを破棄する。

【0051】URLを全て処理したかどうか調べ、処理していなければステップ23に戻って、処理を繰り返す。全て処理していればステップ21に戻る(ステップ28)。以上の処理によって、辞書サーバ中に存在しない語で、今後、翻訳に必要となる可能性が高い語に関する辞書情報

を取得することができる。

【0052】以下では、辞書引き対象語抽出処理(図7、ステップ24)について図10を用いて詳細に説明する。

【0053】辞書引き対象語抽出処理では、サーバ辞書を検索しながら、形態素解析を行うことによって辞書引き対象語を抽出していく。本実施例では、英語を対象とするため、空白を単語の切れ目として採用する単純な方法について述べるが、本発明は例えば、特開昭58-40684号公報に記載されるような形態素解析処理方法についても適用が可能である。

【0054】まず、HTML文書から、HTMLタグを削除し、辞書引き対象語抽出処理の対象となる文字列のみを切り出す(ステップ301)。

【0055】切り出された処理対象文字列を先頭からサーチし、空白文字が見つかったら、直前の空白文字の直後から現在の空白文字の直前までの文字列を単語の候補として切り出す(ステップ302)。

【0056】切り出された単語から、その単語の見出し語の候補の推定を行う(ステップ303)。その際、図11に示すルールを用いる。これは、単語が規則変化する単語である場合には、原形のみが辞書見出しとして登録されているのが一般的であるため、切り出された単語が規則変化する単語の活用形である場合には、切り出された単語は辞書見出しとして登録されていないためである。不規則変換動詞に関しては、活用形についても見出しに登録しておく。

【0057】次に、見出し語の候補が辞書に登録されているかどうかを調べる(ステップ304)。登録されていれば、ステップ307に進む。

【0058】未登録の場合には、全ての見出し語推定ルールを適用したかどうか調べ(ステップ305)、未適用のルールがあれば、ステップ303に戻って見出し語候補の推定の処理を繰り返す。全てのルールが適用済みである場合には、該見出し語候補を辞書引き対象語であると判定し、見出し語を辞書引き対象語として登録する(ステップ306)。

【0059】取得した単語の見出し語を単語テーブル(図示せず)に登録する(ステップ307)。単語テーブルは、現在処理中の文に含まれる単語の見出し語を出現順に格納したものである。

【0060】次に未知語ではない辞書引き対象語の候補を抽出する(ステップ308)。この処理は、後で詳細に説明する(図12、図13)。

【0061】辞書引き対象語が複合語の場合、既に辞書登録されていないかどうかをチェックする。複合語が既に登録されていれば、辞書引き対象語リストから削除する(ステップ309)。

【0062】全てのテキストを処理したかどうかを判定し、未処理のテキストがあれば処理を繰り返す(ステップ310)。

【0063】次に、非未知語辞書引き対象語候補発見処理(図10、ステップ308)を図12を用いて詳細に説明する。

【0064】まず、該見出し語が固有名詞かどうかを判定する(ステップ3081)。単語が文頭以外の位置にある場合に、テキスト中の見出し語の文字列が大文字であれば固有名詞であり、そうでなければ固有名詞ではないと判定する。

【0065】ステップ3081の判定の結果、見出し語が固有名詞である場合には、該見出し語を辞書引き対象語候補として登録する(ステップ3082)。

【0066】次に、該見出し語とその直前に出現した単語を2語単語列として格納する(ステップ3083)。図13は、2語単語列格納テーブルの概念図である。2語単語列格納テーブルは、テキスト中に隣接して出現する2語の単語列の全てを出現頻度と共に格納したものである。2語単語列格納テーブルは、1語目単語リストと2語目単語リストからなる。i語目単語リストは、2語の単語列のi番目に出現した単語の文字列と、i語目までの単語列の出現頻度と、i語目までの単語列に続くi+1単語列のi+1番目の出現単語へのポインタからなる。例えば、図13の場合、1語目単語リストから、Bigが2回出現していることが分かり、Bigのポインタをたどることによって、Bigから始まる2語単語列として、Big Blueというならば、2回出現していることを示している。

【0067】見つかった2語単語列の出現頻度を数え(ステップ3084)、頻度が予め定めた閾値以上であれば、辞書引き対象語の候補として登録する(ステップ3085)。

【0068】次に、図7、ステップ25の他辞書サーバの優先度決定処理について説明する。

【0069】本発明の辞書サーバは、自分のサーバ辞書に見出し語がない場合には他の辞書サーバに見出し語の検索を要求することによりネットワーク上に分散して存在する辞書を利用することができる。このとき、検索対象となる見出し語がどのサーバに存在するか分かっているとき、即ち分散辞書インデックス9に見出し語が存在する場合には、その辞書サーバに検索要求を行う。しかし、見出し語の辞書情報を持つ辞書サーバが分からないとき、即ち分散辞書インデックス9に見出し語が存在しないときには、辞書サーバリスト6中の他の辞書サーバに問い合わせを行う。このとき、辞書サーバリスト6中の任意の辞書サーバに問い合わせを行うと、非常に効率が悪い。また、各辞書サーバが同時に複数の辞書サーバに問い合わせを行うと、指数関数的にトラフィックが増大するために、大きな問題となる。そこで、本実施例の辞書サーバには、翻訳対象である文書を翻訳する際に辞書サーバの優先度を決定する機能を設け、優先度が高い順に問い合わせを行う。辞書サーバの優先度決定処理は、次の3つの方法によって行う。

【0070】第1の方法では、辞書サーバ評価リスト8

を用いて辞書サーバの決定を行う。辞書サーバ評価リスト8の例を図14に示す。辞書サーバ評価リスト8は、辞書サーバリスト6中の各辞書サーバに対して過去に行われた検索要求に対する結果の履歴を格納している。本実施例では、履歴として、検索要求の回数を示す検索頻度、検索が成功した回数を示す成功頻度、検索された訳語がユーザによって変更された回数を示す訳語変更頻度を格納している。ここでは、次の値を各辞書サーバに対して計算し、これが大きい順に検索を行う。

【0071】(評価値) = (成功頻度) \* ((検索頻度) - (訳語変更頻度)) / (検索頻度)<sup>2</sup>

第2の方法では、現在処理対象となっているHTML文書を翻訳するのに特に適切と思われる辞書サーバを決定する。処理対象HTML文書は、検索要求データのURL欄から抽出する。図10のステップ307で、見出し語が存在した語を単語テーブルに登録する際、分散辞書インデックス9で、リモート辞書キャッシュあるいは他辞書サーバに辞書情報が存在することが分かった単語に基づいて最適辞書サーバテーブルを作成する。最適辞書サーバテーブルの例を図15に示す。辞書サーバテーブルは、辞書サーバのIDと各辞書サーバ中に格納されている単語が該HTML文書中に何回出現したかを示している。この最適辞書サーバテーブル中で、出現頻度の高い順に検索を行う。

【0072】第3の方法では、第2の方法と同様に現在処理対象となっているHTML文書を翻訳するのに特に適切と思われる辞書サーバを決定する。辞書中に、その語が使用される分野を格納しておき、ステップ307で、見出し語がサーバ辞書中に存在した語についてその分野を抽出し、テキスト全体の分野を判別する。そして、第1の方法のように、文書の分野毎に辞書サーバに対する検索頻度、成功頻度、訳語頻度を格納しておく。そして、新しい文書に対する辞書情報検索要求を行う場合、まず、文書の分野を判別し、その分野において、最も成功する可能性が高い辞書サーバを選択する。検索が成功する可能性は、第1の方法と同じ式によって計算する。文書全体の分野を判定する方法については、特開平5-257966号公報に記載されているような方法を用いることが可能なので説明は省略する。

【0073】以上、辞書サーバが自動的に未知語を収集する手法を説明した。このような処理により、通信コストを抑えつつ、サーバ辞書に存在しない未知語のうち、今後使用される可能性が高い語の情報を収集することができる。また、辞書サーバの利用者は、自分がアクセスしている辞書サーバのみから、辞書情報を得ることができる可能性が高くなるため、他の辞書サーバを参照しに行くための通信コストが抑えられる。

【0074】(3) 機械翻訳

辞書サーバを用いた機械翻訳方法について説明する。本実施例の機械翻訳方法では、クライアント側で、辞書、

文法を用いて未知語など翻訳不可能部分の認識を行うと同時に、これまで説明してきた辞書サーバに対して辞書情報のような翻訳知識の検索要求を行う。このとき、辞書サーバから辞書情報がすぐに返送されるとは限らない。そこで、クライアント側では、未知語などの翻訳不可能部分に関して翻訳に必要な情報についての仮説を設定し、その仮説に基づいて仮に翻訳を行い、結果を表示する。そして、辞書サーバから辞書情報などの翻訳知識が返送された時点で、返送された翻訳知識に基づいて訳文を変更することにより、クライアント単独で翻訳を行った場合より質の良い翻訳結果を出力する。まず、クライアント側の処理について図16を用いて説明する。以下では、135 is FIFO buffer in which memory access requests are latched.という文を含む文書を翻訳する場合を例に説明する。この例において、latch、FIFOがクライアント辞書14における未知語であると仮定する。

【0075】始めに、クライアント辞書14のみを利用して、入力文を形態素解析する(ステップ401)。この処理は、例えば特開昭58-40684号公報に記載の方法を用いることができるので詳細な説明は省略する。形態素解析は、メモリ13中の形態素解析プログラムによってCPU12で処理される。その際、未知語となった単語は、辞書引き対象語リストに格納する。

【0076】次に、辞書引き対象抽出処理を起動する(ステップ402)。この処理の内容は、図7、ステップ24に示す辞書サーバでの処理と同様なので説明を省略する。例文においては、辞書引き対象語として、未知語であるFIFO、latchが抽出される。

【0077】ステップ402で辞書引き対象語が抽出されると、クライアントは辞書サーバへその辞書引き対象語に関する辞書情報の検索を要求する(ステップ403)。辞書サーバへ送信する検索要求データの構造は、辞書サーバが他の辞書サーバに辞書情報の検索要求を行う際のデータ構造(図8)と同様である。検索要求の対象となった見出し語(例の場合、FIFO、latch)は、検索結果待ちリスト(図示せず)に格納される。クライアントは、未知語の辞書情報を辞書サーバに問い合わせつつ、結果を待つことなく引き続き翻訳対象文書の翻訳処理を進める。

【0078】未知語の翻訳部分に対して、辞書引き対象語仮説設定処理を行う(ステップ404)。この処理は、辞書サーバへの辞書引き対象となった語を含む文の翻訳処理を辞書サーバへの検索と非同期に実行するためのものであり、メモリ13中の仮説設定プログラムによって実行される。この処理では、形態素解析の結果として得られる単語テーブルにおいて、辞書引き対象語については品詞を推定した結果を格納し、その単語テーブルを用いて翻訳を実行する。品詞の仮説設定ルールとしては、図17に示すようなものを用いる。仮説設定処理は、図11に示した見出し語推定ルールと図17の仮説設定ルールの組み合わせによって行う。即ち、見出し語推定ルール

によって、見出し語を推定し、その見出し語に対し仮説設定ルールが適用できるかどうかを調べるというように処理を行う。例の場合、FIFOの場合は、図17の副詞、形容詞、動詞を示すルール(条件)にマッチしないため、それ以外という条件から、名詞という仮説が設定される。同様な処理により、latchedの場合は、動詞の過去形という仮説が設定される。

【0079】仮説が設定されると、その仮説を利用して構文解析(ステップ405)、構文変換(ステップ406)、生成処理(ステップ407)を実行し、訳文を得る。具体的には、メモリ13中の構文解析プログラム、構文変換プログラム、生成プログラムによってCPU12で処理が実行される。これらの処理方法は、例えば、特開昭58-40684号公報に記載の方法を用いることができるので詳細な説明は省略する。

【0080】この時点では、FIFO、latchは、未知語であるため、完全な訳文は得られていないため、図18に示すように翻訳前の単語を含む訳文が表示される(ステップ408)。図18において、FIFO、latchの部分をクリックにしても良い。また、辞書サーバへの辞書検索要求中であることを示すために、強調して表示することも効果がある。

【0081】次に、仮説記録ファイルを作成する(ステップ409)。仮説記録ファイルは、翻訳対象文中に辞書引き対象語が存在した場合に、辞書引き対象語に関して立てた仮説を格納するファイルである。仮説記録ファイルの例を図19に示す。仮説記録ファイルは、辞書引き対象語毎に、その語が出現した文の原文ファイル中での位置、翻訳結果の文の訳文ファイル中での位置、辞書引き対象語の訳語の位置、仮定された品詞、などを各出現位置毎に格納している。図19の例は、次のようなことを示している。

【0082】入力文 135 is FIFO buffer in which memory access requests are latched.は、翻訳対象文書中で、115バイト目から始まり、176バイト目で終わっており、訳文「135は、メモリアクセス要求がlatchされるFIFOバッファである。」は、訳文文書中の120バイト目から始まり、177バイト目で終わっている。また、未知語latchの訳語「latchされる」が、147バイト目から始まり、151バイト目で終わっており、未知語FIFOの訳語「FIFO」が、158バイト目から始まり、161バイト目で終わっている。

【0083】次に、ステップ403で辞書サーバに対して送信した検索要求に対し、辞書サーバから検索結果を受信したかどうかを調べる(ステップ410)。検索結果は、翻訳処理とは非同期に受信され、テンポラリ辞書(図示せず)に格納される。検索結果待ちリスト中の語をテンポラリ辞書で検索し、見出し語として格納済みであれば、検索結果が受信済みであると判定する。検索結果の構造は、辞書サーバが他の辞書サーバに送った検索要求

に対して受信した結果のデータ構造(図9)と同様である。ただし、内容種別は、常にCであり、辞書情報が格納されているものとする。暗号化フラグが0の場合には、そのまま辞書情報をテンポラリ辞書へ複写する。暗号化フラグが1の場合には、暗号鍵テーブルを参照して、該辞書サーバの辞書サーバIDから暗号鍵を取得して、復号処理を行い、復号処理後の情報をテンポラリ辞書に複写する。図20に、暗号鍵テーブルの例を示す。暗号鍵テーブルは、辞書サーバIDと該辞書サーバからの辞書情報を復号するのに必要な暗号鍵のリストである。暗号鍵は、クライアントのユーザが予め各辞書サーバの管理者と契約することによって取得しておく。これにより、各辞書サーバでは、契約をしたクライアントのユーザに対してのみ辞書情報を提供することが可能となる。

【0084】検索結果を受信済みならば、メモリ13中の仮説検定プログラムによって仮説記録ファイルを参照し、検索結果と仮説の比較を行い、さらに、検索結果待ちリストから該見出し語を削除する(ステップ411)。検索結果を受信していなければ、次の文に進んで処理を継続する。仮説が正しければ、仮説記録ファイルから、辞書引き対象語の訳語の文字列位置を取得し、検索した訳語に対し、仮説記録ファイル中の属性に基づいて語形変化を施したものと置き換える(ステップ412)。もし、仮説が誤っていたら、原文の文字列を取得し、検索した辞書情報を用いて再度翻訳を行い、訳文の文字列全体を置き換える(ステップ413)。そして、変更された訳文を表示する(ステップ414)。訳文の表示は、図2.1のようになる。

【0085】次に、サーバ側の処理について図22を用いて説明する。

【0086】クライアント側の処理(図16、ステップ403)において、ある見出し語の検索がサーバに要求されると、辞書サーバは、分散辞書インデックス9を検索して、該見出し語が存在するかどうかを調べる(ステップ101)。存在しない場合には、ステップ109に進む。存在した場合には、ステップ102に進み、対応する辞書情報ポインタを取得して(ステップ102)、ポインタの種別を調べ、サーバ辞書を指していればステップ104へ、リモート辞書キャッシュを指していればステップ105へ、他辞書サーバを指していればステップ106へ進む(ステップ103)。

【0087】ポインタ種別がサーバ辞書を指すとき、サーバ辞書から辞書情報を読み出す(ステップ104)。ただし、検索要求データ中のURL欄の値に対応するURL対応辞書が存在する場合には、URL対応辞書に見出し語が存在するかどうかをチェックし、存在した場合にはURL対応辞書中の辞書情報を読み出す。

【0088】ポインタ種別がリモート辞書キャッシュを指すとき、リモート辞書キャッシュから辞書情報を読み出す(ステップ105)。見出し語が存在し、ポインタ種別

が他の辞書サーバを指すとき、他辞書サーバへ検索対象種別Lで、辞書検索要求を行う(ステップ106)。ここで、検索対象種別をLとするのは、検索要求を行う他辞書サーバが辞書情報を持っていることが判明しているためである。それから、他辞書サーバへの検索要求が成功したかどうか調べる(ステップ107)。成功していればステップ108の処理を行い、失敗していたら、ステップ109の処理を行う。

【0089】得られた検索結果を検索要求元のクライアントに返送する(ステップ108)。返送データの構造は図9に示したのと同様である。また、ここでは、辞書サーバでの検索要求が再帰的に行われた際、得られた検索結果を元の検索要求を行ったクライアントに直接返送する方法について説明したが、それ以外にも、検索要求を仲介した辞書サーバを介して返送することも可能である。

【0090】ステップ106において他の辞書サーバでの検索が失敗した場合は、検索対象種別がRであるかどうか調べる(ステップ109)。Rであれば、ステップ110に進む。Rでなければ、ステップ113に進む。

【0091】検索対象種別がRの場合には、他の辞書サーバへの辞書検索を打ち切るかどうかを判定し、打ち切らない場合には、他の辞書サーバへ、該見出し語の辞書情報検索を依頼する(ステップ110)。辞書検索打ち切りの判定は、検索要求データ中の検索深さおよび検索時刻に基づいて行う。検索対象辞書サーバの決定は、ステップ25で述べた優先度に基づいて行う。詳細は、ステップ25で述べたので説明を省略する。そして、検索要求が成功したかどうか調べ(ステップ111)、成功していれば辞書サーバ評価リスト8を更新する(ステップ112)。辞書サーバ評価リスト8は、図14に示したのと同様である。ユーザによる訳語の変更については、後で説明する。ここでは、該当する辞書サーバの検索頻度を1回増し、成功頻度を1回増す。

【0092】ステップ109で検索種別がRでないとき、及び、ステップ110での検索要求が失敗のときには、辞書サーバ評価リスト8を更新する(ステップ113)。ここでは、検索頻度のみを1回増す。そして、エラーコードを検索要求元に返送する(ステップ114)。最後にURL翻訳頻度リスト7を更新し(ステップ115)、処理を終了する。

【0093】以上の処理によって、クライアントから辞書情報の検索要求があった見出し語に関する辞書情報をクライアントに返送することができる。

【0094】本実施例におけるクライアントは、翻訳専用の装置である必要はない。情報処理機能を有する一般的な情報処理装置に通信機能や翻訳機能を付加するコンピュータプログラムをインストールして同等の処理を行うことができる。

【0095】通信機能を負荷する通信プログラムにより、ネットワークを介して接続された他の情報処理装置に検索を要求したり、何らかの情報を受信することが可

能になる。翻訳の場合は、語句の意味、文法上の役割、派生語やその語句が生まれた背景等様々な情報を受信することが考えら得る。

【0096】翻訳機能を付加するプログラムは以下の処理を指示するものである。

【0097】まず、翻訳すべき翻訳対象文書を形態素解析し（図16、ステップ401に対応）、辞書引き対象語句の意味を自端末の保有する辞書で調べる（図16、ステップ402、403に対応）。辞書は、翻訳プログラムの一部として準備されて翻訳プログラムとともにクライアント側の情報処理装置にインストールされても良いし、所望の言語の辞書データベースを別個に準備しても良い。いずれにしても辞書はクライアント側装置として閉じる範囲に設定され、辞書引きの対象が容易に実行できるようにすることが必要である。

【0098】クライアント側装置に備えられた辞書を用いた辞書引きが成功した場合は、辞書から得られた訳語を用いて、構文解析、構文変換、訳文生成、訳文表示（図16、ステップ405～408に対応）を行い、翻訳単位（文章、文節、文法的、意味的等の所定のまとまり）の翻訳処理を終了し、次の翻訳単位にも同様の処理を行い次々と翻訳文をクライアント側の情報処理端末の表示装置に表示していく。

【0099】翻訳処理中に、クライアント側の端末の辞書に語句（辞書引き語）の訳語が見つからない場合には、その訳語（未知語）に仮説（又は仮訳）を設定する（図16、ステップ404に対応）。

【0100】そして、クライアント側の装置の処理としては、その仮説（又は仮訳）が正しいものとして、引き続き構文解析～訳文表示（但し仮説（又は仮訳）部分は下線、文字色や背景色を変えたり翻訳前の語句のまま表示する）を行う。

【0101】一方、仮説を設定された語句は、上記のようにクライアント側装置での翻訳確定を一旦停止する。そして他の後続部分の翻訳が引き続き行われるが、その後続部分の翻訳処理が行われている間に、クライアント側の装置はネットワークを介して他の情報処理装置に未知語に関する情報の検索を依頼する（図16、ステップ403に対応）。

【0102】未知語に関連する情報は、辞書における未知語の意味であったり、又は文法的なルールであるなど、とにかく、翻訳に必要な情報で、かつクライアント側の装置だけでは知り得ない情報を意味する。

【0103】クライアント側装置では他の情報処理装置に問い合わせ中の未知語に対して設定した仮説をテーブル化して管理する（図16、ステップ409に対応）。

【0104】クライアント側装置は未知語の翻訳に仮説を設定したまま途中で中断することなく文書の翻訳を続けているが、他の情報処理装置へ送信した未知語に関する情報の検索については、結果の受信を監視する。検索

結果を受信した場合には（図16、ステップ410に対応）、その未知語の翻訳に対して設定した仮説と受信した検索結果を比較し、その結果、仮説が誤っていたときには得られた情報を元にして翻訳をする。仮説が正しかったときには、その仮説に基づいた翻訳を確定し（図16、ステップ411～413に対応）。そして、最終的な訳文表示の修正等をして翻訳対象文書の翻訳が完了するまで、翻訳単位の読み込み、仮説設定要否、仮説問い合わせと翻訳の同時進行、仮説に関する情報の入手判定と翻訳結果との併合が行われる。

【0105】以上の処理を行うプログラムは、フロッピーディスク等の磁気記録媒体、CD-ROM、DVDディスク、光カード等の光記録媒体、光磁気記録媒体、フラッシュメモリ、ICカード等の半導体メモリに保持され、コンピュータシステムのCPUに読み込まれたり、ネットワークを介して他の情報処理装置の記憶装置から自端末のメモリにダウンロードしてCPUで処理を行わせることが可能である。

【0106】なお、本実施例では、翻訳不可能部分が未知語である場合に、未知語の辞書情報を検索する場合を例として説明したが、本発明はそれ以外の翻訳知識の検索にも容易に拡張できる。例として、構文解析のための共起データを翻訳知識サーバから検索する場合について説明する。

【0107】共起とは、同時に出現しやすい意味的に関連のある語の組であり、このような共起データを蓄積することによって機械翻訳処理における曖昧性を解消することができる。図23に例を示す。図23の例文は、構文的に2つの曖昧性を持っている。即ち、前置詞句with four doorsが、a carという名詞句を修飾しているのか、それとも、bought a carという動詞句を修飾しているかである。この場合、(car, door)という共起(車にはドアがあることを示す)を格納しておくことにより、正解を得ることができる。

【0108】このような共起データは、語の組の情報であることから、辞書情報の2乗の数が必要であり、辞書よりも大規模なものになる。そのため、共起情報をユーザが個別に持つことは現実的ではない。

【0109】本発明をこのような場合に適用すると、翻訳処理における構文解析時に曖昧性が発生した時点で、共起情報の検索を辞書サーバに要求し、それと同時にデフォルト情報(通常、最も近い係り先を修飾すると判定する)に従って翻訳処理を仮に実行して結果を出力し、共起情報が返送された時点で、デフォルトの解析結果と共起情報による解析結果を比較し、正しければそのまま処理を続行し、正しくない場合には、共起情報を用いて解析処理を実行することによって正しい訳文を得ることができる。

【0110】さらに、クライアント側で翻訳不可能部分を認識した場合には、翻訳不可能部分を含む文を辞書サ

サーバに送信し、辞書サーバにおいて翻訳処理を実行してその結果を返送し、クライアントによる結果とサーバによる結果が一致する場合には、処理をそのまま実行し、異なる場合には、サーバによる翻訳結果で置換する構成とすることも可能である。

【0111】以上述べたように、本発明は辞書情報のみに限定して適用されるものではなく、クライアントと翻訳知識を提供する上位のサーバとの組み合わせにより、ネットワークを介した高品質な翻訳を提供するものである。

#### 【0112】(4) 辞書情報の共有

次に、辞書サーバを用いた機械翻訳方法におけるユーザによる訳語変更情報共有処理について説明する。対訳辞書では、一つの見出し語に対して、複数の訳語が存在する場合がある。この場合、翻訳時に機械翻訳システムが選択した訳語が誤っている場合がある。このようなときに、ユーザの指示に従って他の訳語を表示し、ユーザに選択させるシステムが従来知られている。さらに、このようなユーザによる選択情報を学習することによって、それ以降の訳語選択処理を正しく行うための方法が提案されている。例えば、特開平7-6175号公報には、後編集における訳語選択時に新しい訳語を自動的にユーザ辞書に格納する技術が開示されている。このような技術は、ユーザが翻訳を希望するある特定の分野では、訳語が一意に決定できる可能性が高いという考え方に基づいている。しかし、従来の方法では、高い翻訳精度が得られない。例えば、presidentという語は、「社長」、「大統領」という2つの訳語を持つ。しかし、分野のような大雑把な捉え方では、presidentのような単語の訳語を正しく決定することはできない。

【0113】本発明では、辞書サーバを多くのクライアントが共用する構成としたことで、同一のHTML文書に關する辞書情報の検索が多くの回数行われることに着目し、ユーザによる訳語選択情報を文書毎に保存することによって、高い精度で訳語選択を行えるようにする。以下、訳語変更情報共有処理について説明する。

【0114】まず、クライアント側の処理について図24の処理フローを用いて説明する。以下、図25で示すように、presidentの訳語が誤って「大統領」になった例を用いて説明する。

【0115】ユーザは、訳文表示を読んでいるときに訳語が誤っていたり、意味が疑わしい箇所を見つけると画面上の該当箇所をマウスのようなポインティングデバイスで指示する。例の場合、ユーザは、presidentを指示する。このとき、システムは、ユーザが表示画面上で指示した単語を検出する(ステップ501)。なお、ユーザが「大統領」を指示した場合に、日英の辞書や翻訳処理によって得られる英語単語と日本語単語の対応関係を参照して、presidentを得るようにシステムを拡張することは容易である。次にその単語を見出し語として、辞書を

検索し、見出し語があるかどうか調べる(ステップ502)。この際、変化形の処理については、ステップ303に述べた通りの処理を行う。また、ここで辞書検索の対象となるのは、辞書サーバから検索したテンポラリ辞書とクライアントが持っている辞書の両方である。もし、この中に見出し語があれば、訳語の候補を全て表示する(ステップ504)。訳語の候補の表示例を図26に示す。もし、辞書中に見出し語が見つからなければ、その旨をユーザに表示した後、ユーザによる単語指定を待つ。

10 【0116】訳語候補が表示された場合には、ユーザは、訳語候補の表示を見て、正しい訳語があれば、それをマウスなどで指示する。システムは、ユーザが訳語を変更を指示したかどうかを検出し(ステップ505)、訳語の変更が指示されたら、訳語を変更して表示する(ステップ506)。これにより、訳文は、図27に示すように変更される。最後に、システムは、ユーザが指定した見出し語と選択した訳語の組を辞書サーバに送信する(ステップ507)。

20 【0117】次に、訳語変更情報共有処理における辞書サーバ側の処理を図28を用いて説明する。辞書サーバは、訳語変更情報を受信すると、受信したデータ中の見出し語を分散辞書インデックス9で検索する。該見出し語がなければステップ607に進む。あれば、辞書情報ポインタを取得し(ステップ602)、ポインタの種別を判定して、Cならばステップ604へ、Rならばステップ605へ、Lならばステップ607に進む(ステップ603)。リモート辞書キャッシュから辞書サーバ名を取得する(ステップ604)。分散辞書インデックス9から辞書サーバ名を取得する(ステップ605)。取得した辞書サーバ名を辞書サーバ

30 評価リスト8中で検索し、訳語変更頻度を1回増す(ステップ606)。受信したデータから、URL対応辞書を作成する(ステップ607)。URL対応辞書は、図29のような構成になる。なお、ここでは、URL対応辞書を別ファイルとして構成したが、サーバ辞書、リモート辞書キャッシュ中に訳語を登録し、訳語選択条件としてURLを記述するような構成としても同様の効果を得ることができる。

40 【0118】以上のようにクライアント側でのユーザによる訳語変更をURL対応辞書に反映することにより、あるユーザがあるURLに対して選択した正しい訳語情報を他のユーザが共用することが可能になる。なお、本発明の実施例では、クライアント側の検索要求に応じて辞書情報を送付する構成としたが、URL対応辞書中の辞書情報は、クライアント側での未知語以外の語の翻訳に対しても有用である。そのため、クライアントでの翻訳開始時点において、未知語の有無に関わらず翻訳対象文書のURLを辞書サーバに送付し、翻訳対象文書のURLに対応する辞書が存在する場合には、URL対応辞書中の辞書情報をまとめてクライアントに送付する構成を取ることができ

50 【0119】以上の処理によって、ユーザは、辞書サー

バ上にある大規模な辞書を利用することができ、ローカルにクライアントが持つ辞書のみを利用する場合と比較して、良い品質の訳文を得ることができる。さらに、ネットワークのスピードが遅い場合にも、辞書情報が得られた順に訳語を表示することができるため、全ての情報が得られてから処理を行う場合と比べて、ユーザの使い勝手が良い。また、ユーザが訳語を変更した情報をサーバ側に返送し、その情報を共用することにより、より良い品質の訳語を得ることができる。さらに、辞書サーバは、クライアントから検索要求が来る可能性が高い語を予め収集しておくことによって、ユーザからの検索要求に高速に応答することができ、さらに、辞書サーバ上に辞書情報がない場合にも、要求された見出し語を持つ可能性が高い他の辞書サーバを推定することにより、少ない通信コストで所望の辞書情報をクライアントに返送することが可能となる。

## 【0120】

【発明の効果】辞書サーバは、サーバ辞書とサーバ辞書中の単語および他の辞書サーバ中に存在する辞書の見出しを格納する分散辞書インデックスを備え、クライアントからの検索要求に対し、サーバ辞書中に存在する語については、サーバ辞書中の辞書情報を返送し、他の辞書サーバ中に存在する語については、他の辞書サーバに検索要求を行う。さらに、どの辞書サーバに検索対象となった語が存在するかが分からない場合には、辞書情報を持つ可能性が高いサーバを統計的に推定し、検索要求を行う。これにより、ネットワーク上に分散する辞書の辞書情報検索を効率的に行うことが可能となる。

【0121】クライアントは、クライアントにおいて翻訳処理を実行中に未知語などの翻訳不可能部分を認識し、上記翻訳不可能部分を含む構文的にまとまりのある構文単位についての翻訳知識の検索を辞書サーバに要求し、上記構文単位に続く内容の翻訳を続行し、辞書サーバから翻訳知識が返送された時点で、上記翻訳不可能部分を含む翻訳対象文書全体の訳文を作成する。これにより、ネットワークの速度が遅く、辞書サーバへの検索要求に要する時間が大きい場合でも、ユーザにとって使い勝手が良い翻訳システムを提供できる。

【0122】また、ユーザによる訳語変更情報を翻訳対象となったHTML文書のURLに対応して格納することによ

10

20

30

＊って、他のユーザが同じ文書を再翻訳する際に、より品質の高い訳語を提供することが可能となる。

## 【図面の簡単な説明】

【図1】本発明の実施例の構成図である。

【図2】URL翻訳頻度リストの概念図である。

【図3】分散辞書インデックスの概念図である。

【図4】新規登録語情報送信処理の処理フローである。

【図5】新規登録語情報の概念図である。

【図6】新規登録語情報受信処理の処理フローである。

【図7】辞書情報収集処理の処理フローである。

【図8】検索要求データの概念図である。

【図9】検索結果返送データの概念図である。

【図10】辞書引き対象語抽出処理の処理フローである。

【図11】見出し語推定ルールの例である。

【図12】辞書引き対象語候補発見処理の処理フローである。

【図13】2語単語列保持用データの概念図である。

【図14】辞書サーバ評価リストの概念図である。

【図15】最適辞書サーバテーブルの概念図である。

【図16】翻訳処理におけるクライアント側の処理の処理フローである。

【図17】品詞推定ルールの例である。

【図18】未知部分を含む表示画面例を示す図。

【図19】仮説記録ファイルの概念図である。

【図20】暗号鍵テーブルの概念図である。

【図21】訳文の表示画面例を示す図。

【図22】翻訳処理におけるサーバ側の処理の処理フローである。

【図23】構文的曖昧性の例である。

【図24】URL対応辞書作成処理におけるクライアント側の処理の処理フローである。

【図25】システムによって選択された訳語が誤っている訳文の表示例である。

【図26】訳語候補の表示例である。

【図27】訳語変更後の訳文の表示例である。

【図28】URL対応辞書作成処理におけるサーバ側の処理の処理フローである。

【図29】URL対応辞書の概念図である。

【図2】

図2

	URL	x月y日	...	...	...
1	http://info.../robots.html	60	200	100	40
2	http://net...com/index.html	0	0	60	50
3	http://www...com	290	400	1600	1600
	...	...	...	...	...

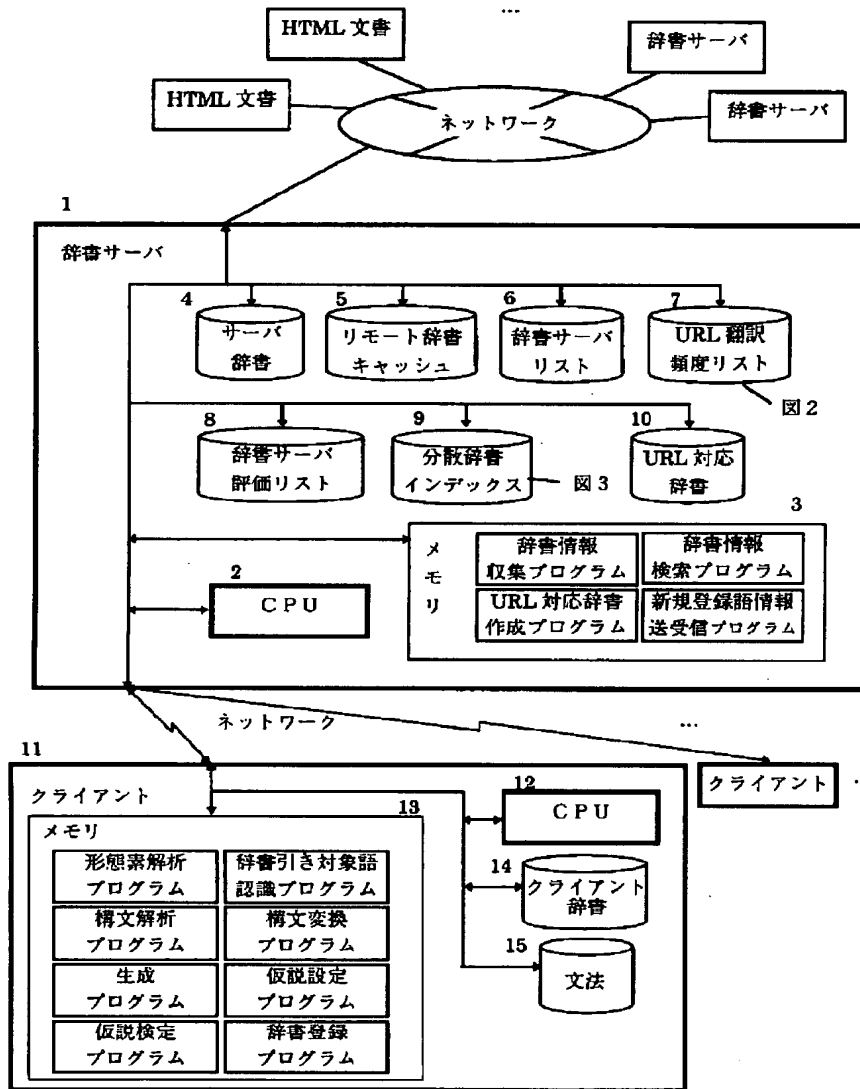
【図3】

図3

見出し	辞書情報ポイント	辞書情報ポイント	辞書情報ポイント
	ポイント値	ポイント値	ポイント値
Big Apple	L 16435		
Big Blue	C 44825	R 133.144.18.152	
Big Dipper	R 133.144.18.150	R 133.144.18.154	R 133.144.18.164
...	...	...	...

【図1】

図1



【図5】

図5

辞書サーバID	見出し語1	見出し語2	見出し語3	...
---------	-------	-------	-------	-----

【図15】

図15

辞書サーバ名	出現頻度
辞書サーバ1	80
辞書サーバ2	40
辞書サーバ3	10
...	...

【図17】

図17

条件	品詞
最後の文字列が"ly"	副詞
最後の文字列が"ive"	形容詞
最後の文字列が"ate"	動詞
最後の文字列が"ize"	動詞
最後の文字列が"ed"	動詞、形容詞
...	...
それ以外	名詞

【図20】

図20

辞書サーバ名	暗号鍵
辞書サーバ1	102340923789248
辞書サーバ2	192873921372395
...	...

【図8】

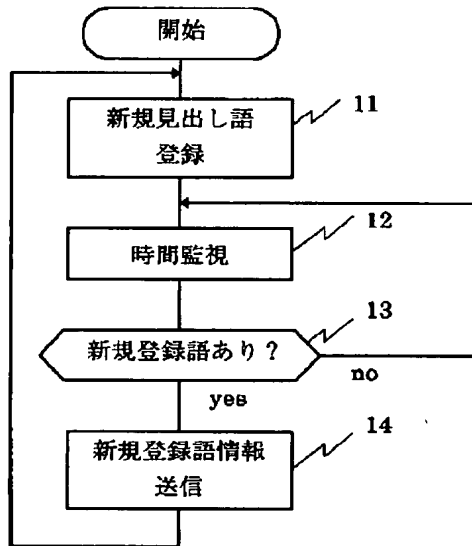
図8

検索対象種別	検索深さ	見出し語	URL	検索時刻
--------	------	------	-----	------



【図4】

図4



【図9】

図9

要求元	辞書サーバ名	成功フラグ	内容種別	暗号化フラグ	辞書情報
-----	--------	-------	------	--------	------

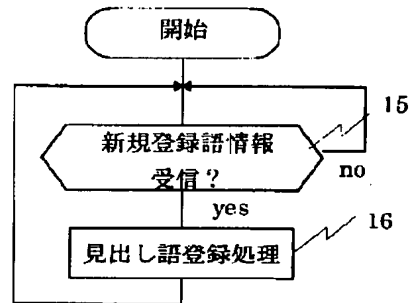
【図11】

図11

条件	見出し語文字列生成方法
なし	なし
末尾が ves	ves を f または fe に変える。
末尾が ine	ine を y に変える。
末尾が es	es を e に変える。
末尾が s	s を取る。
末尾が ind	ied を y に変える。
末尾が子音2個+ed	ed と子音1個を取る。
末尾が ed	ed を取る。
末尾が d	d を取る。
末尾が子音2個+ing	ing と子音1個を取る。
末尾が ing	ing を e に変える。
末尾が ing	ing を取る。

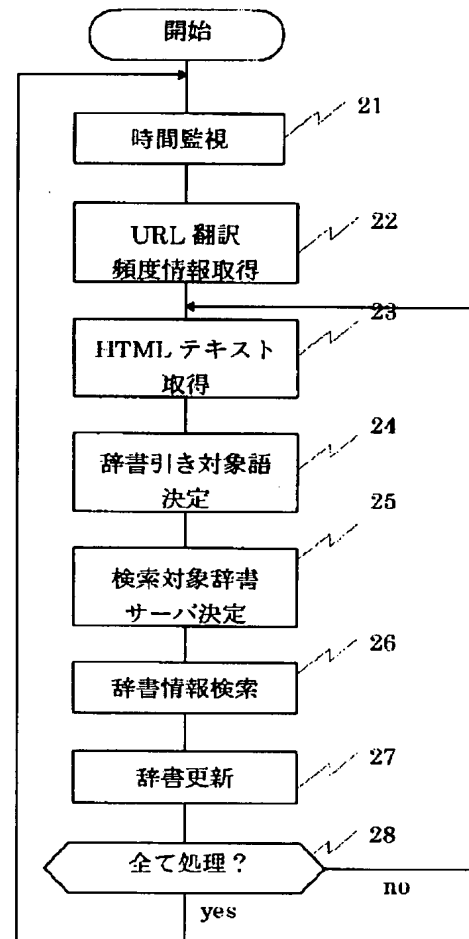
【図6】

図6

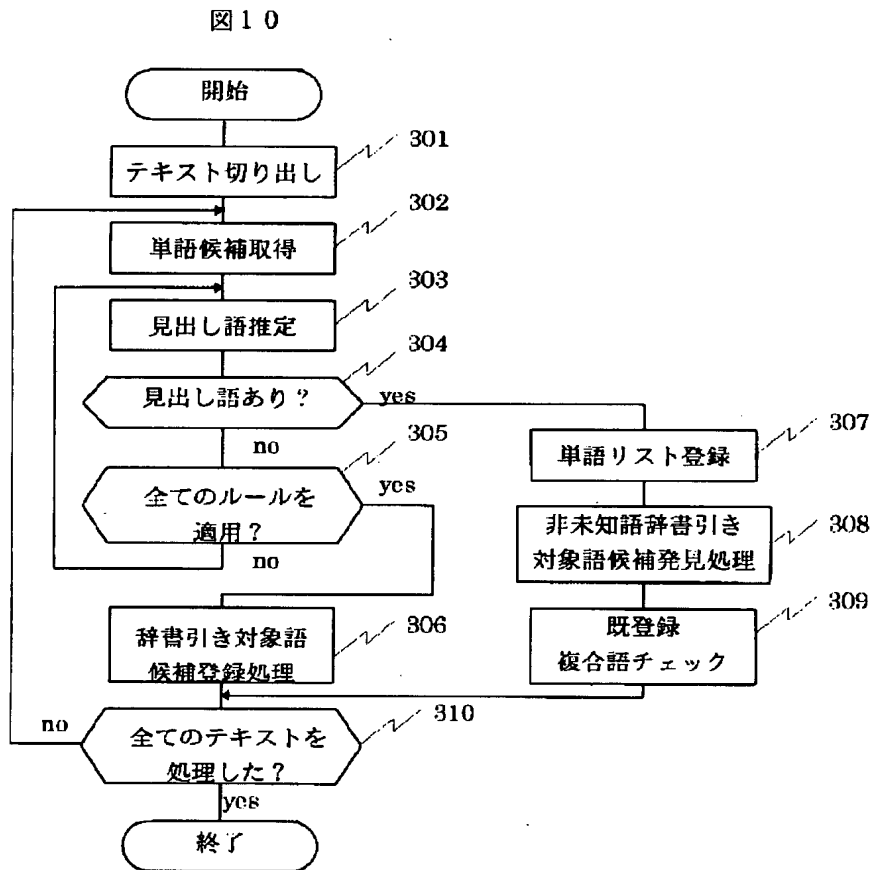


【図7】

図7



【図10】



【図13】

図13

1 語目リスト		2 語目リスト	
Big	2	Blue	2
Blue	3	buy	1
buy	1	Lotus	1
Lotus	1	announce	1
...	1	...	1

【図14】

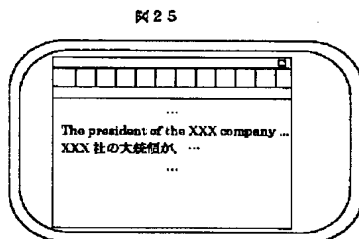
図14

辞書サーバIPアドレス	検索頻度	成功頻度	訳語変更頻度
188.144.18.150	100	18	2
193.144.18.154	200	160	40
193.144.18.97	248	200	10
...	...	...	...

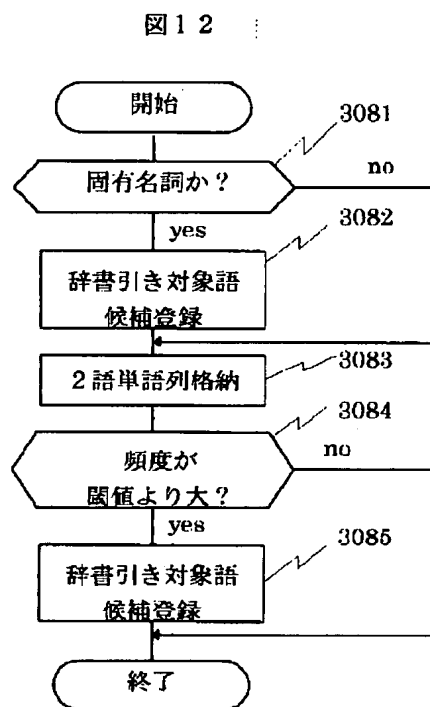
【図18】



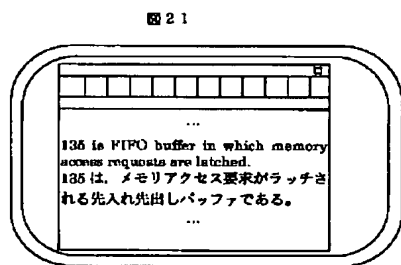
【図25】



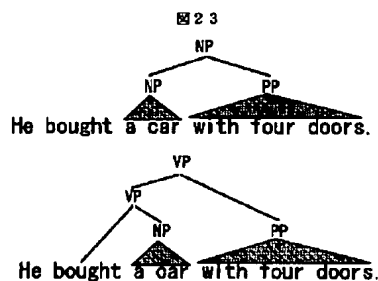
【図12】



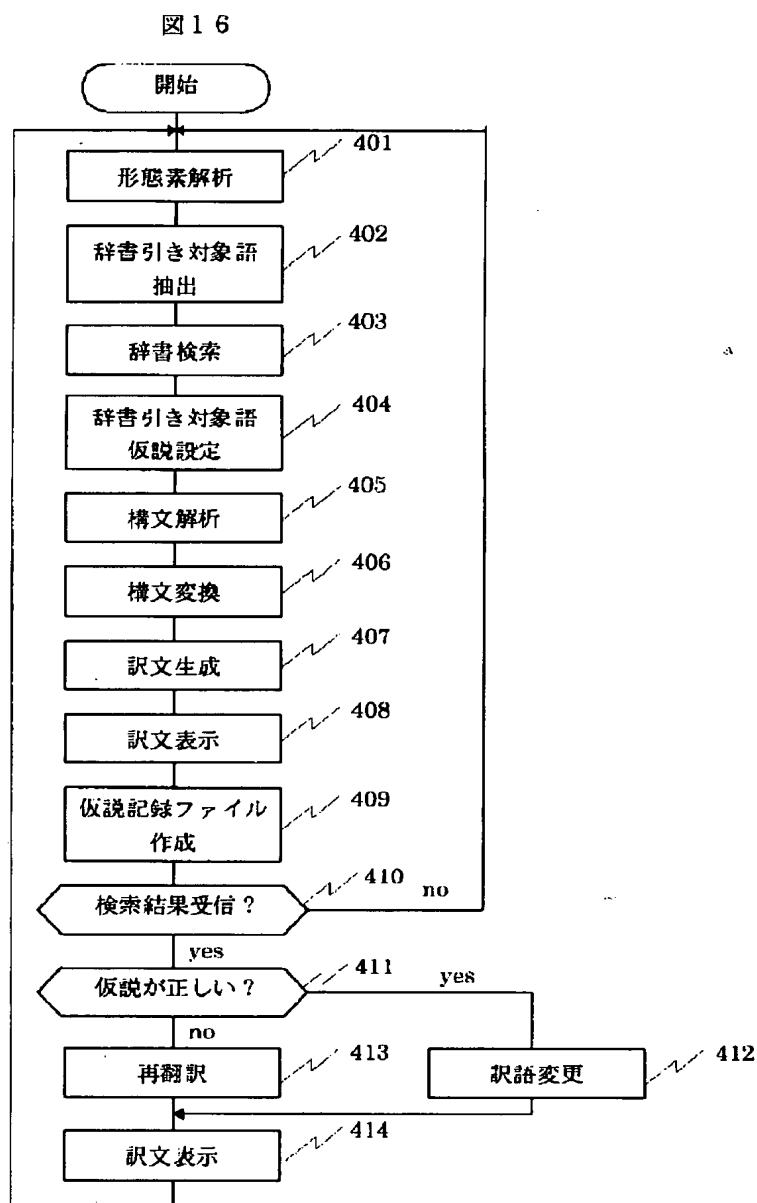
【図21】



【図23】

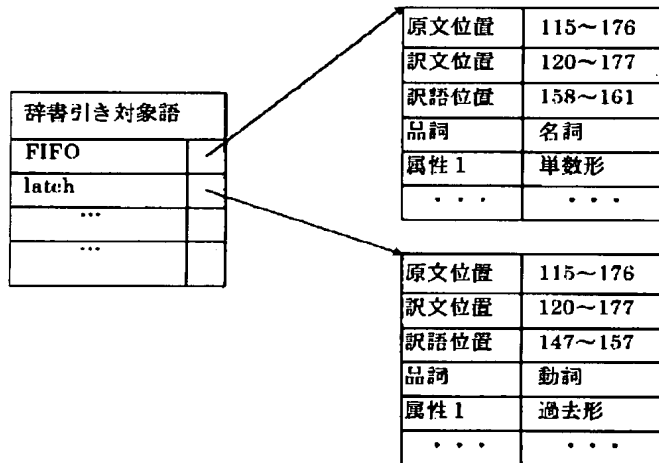


【図16】



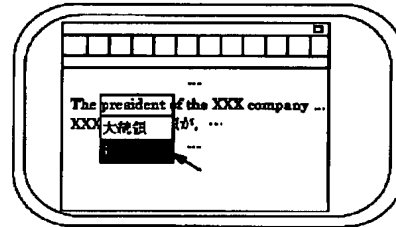
【図19】

図19



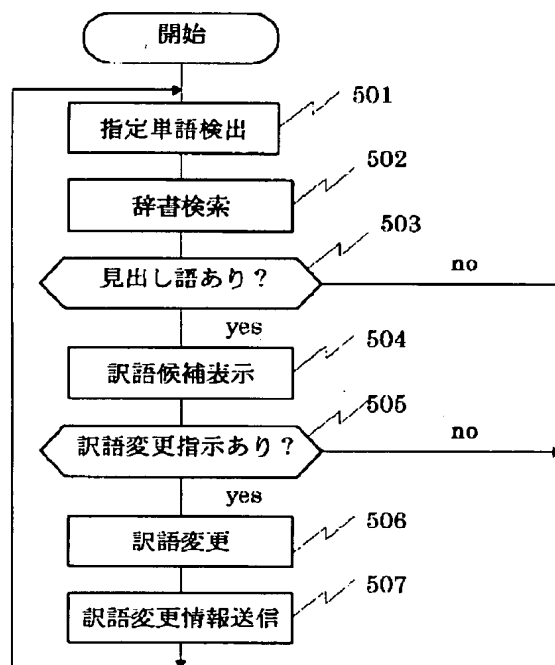
【図26】

図26



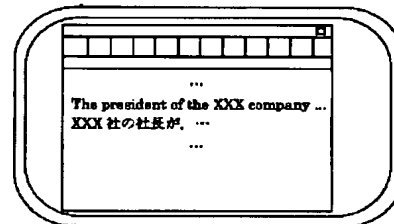
【図24】

図24



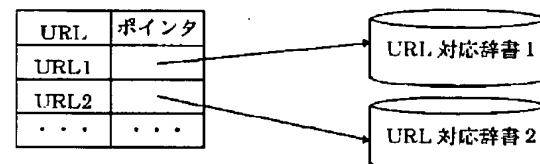
【図27】

図27



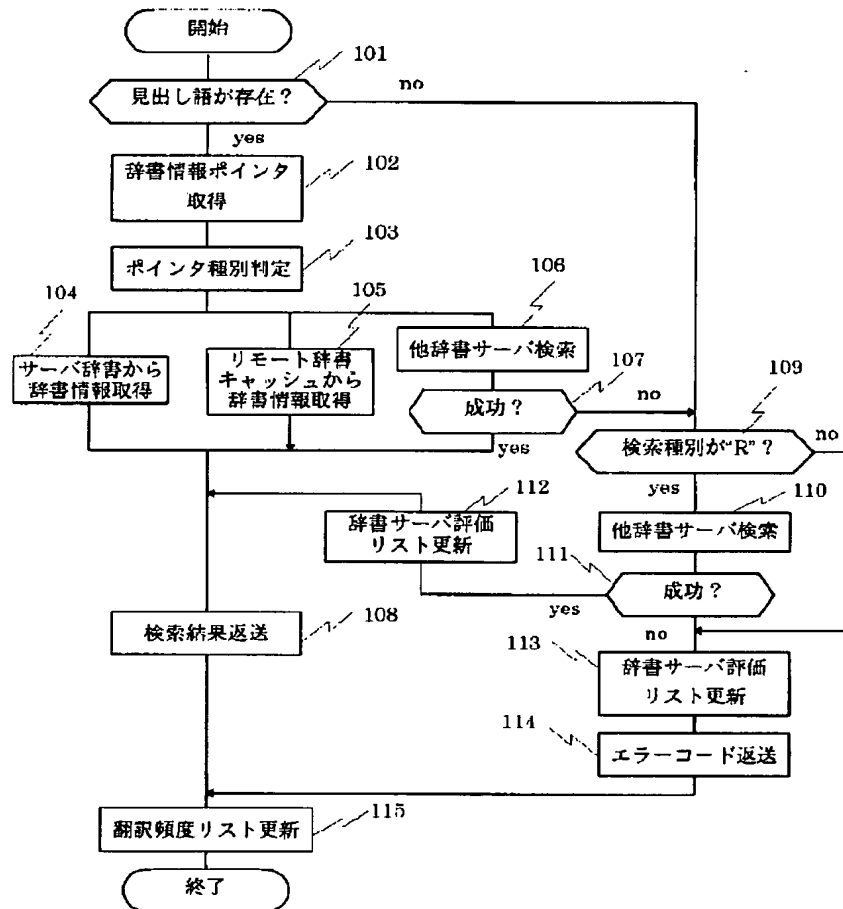
【図29】

図29



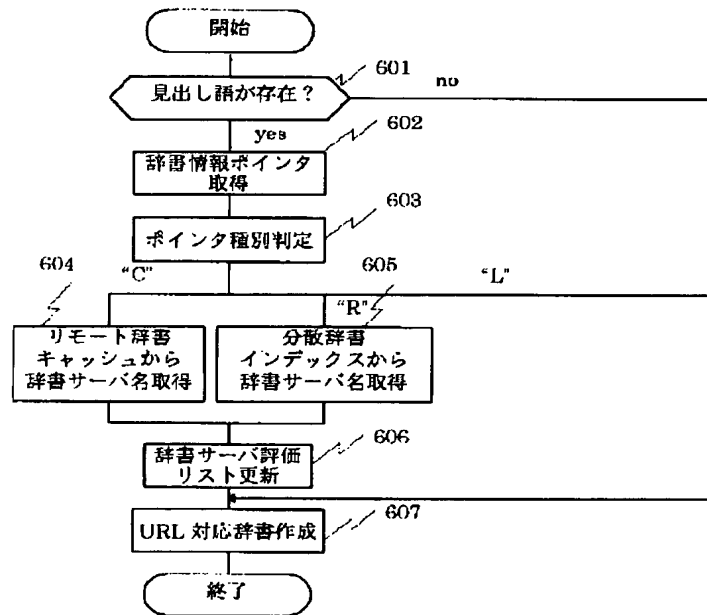
【図22】

図22



【図28】

図28



【公報種別】特許法第17条の2の規定による補正の掲載  
 【部門区分】第6部門第3区分  
 【発行日】平成14年7月12日(2002.7.12)

【公開番号】特開平10-198680  
 【公開日】平成10年7月31日(1998.7.31)  
 【年通号数】公開特許公報10-1987  
 【出願番号】特願平9-547  
 【国際特許分類第7版】

G06F 17/28  
 17/30

【F I】

G06F 15/38 C  
 15/40 310 C  
 370 J

【手続補正書】  
 【提出日】平成14年4月5日(2002.4.5)  
 【手続補正1】  
 【補正対象書類名】明細書  
 【補正対象項目名】発明の名称  
 【補正方法】変更  
 【補正内容】  
 【発明の名称】 機械翻訳方法、機械翻訳結果の表示方法、分散辞書検索方法及びネットワーク分散型機械翻訳システム及びそれを実現するためのコンピュータ可読媒体  
 【手続補正2】  
 【補正対象書類名】明細書  
 【補正対象項目名】特許請求の範囲  
 【補正方法】変更  
 【補正内容】  
 【特許請求の範囲】  
 【請求項1】ネットワークに接続され、翻訳知識を保持する記憶手段を有する情報処理装置を用いて実行される機械翻訳方法であって、上記翻訳知識を参照して翻訳対象文書を翻訳する際に、上記翻訳対象文書中で上記翻訳知識では翻訳できない翻訳不可能部分を認識し、少なくとも該翻訳不可能部分を含み、文法的にまとまりのある構文単位に関する翻訳知識を該ネットワークに接続された他の情報処理装置に問い合わせ、上記構文単位に続く内容の翻訳を続行し、上記他の情報処理装置から上記翻訳不可能部分に関する翻訳知識を得て、上記翻訳不可能部分を含む上記翻訳対象文書の翻訳結果を作成することを特徴とする機械翻訳方法。  
 【請求項2】上記情報処理装置は上記翻訳不可能部分を含む構文単位について仮の翻訳処理を行った結果を出力し、上記他の情報処理装置から得られた翻訳知識に基づいて翻訳結果を作成する処理では、該仮翻訳結果を上記翻訳知識に基づいて変更することを特徴とする請求項1

の機械翻訳方法。

【請求項3】上記仮の翻訳処理では、上記翻訳対象文書における未知語の品詞を推定し、該推定された品詞に従って翻訳処理を実行し、翻訳結果を変更する処理では、該推定された品詞と他の情報処理装置から得られた該未知語の品詞を比較し、一致していれば訳語のみを変更し、一致していなければ該未知語を含む文全体を再翻訳することを特徴とする請求項2に記載の機械翻訳方法。

【請求項4】機械翻訳装置における翻訳結果の表示方法であって、翻訳対象文書を翻訳する際に、その時点で利用可能な翻訳知識を用いて翻訳を実行して、翻訳結果を出力し、利用可能な翻訳知識に変更があったかどうかを認識し、変更があった場合には、該変更に基づいて過去の翻訳結果の出力を変更することを特徴とする機械翻訳結果の表示方法。

【請求項5】ネットワークに接続され、辞書と文法を保持する記憶手段を有する情報処理装置を用いて実行される機械翻訳方法であって、上記辞書と文法を参照して翻訳対象文書を翻訳する際に、上記翻訳対象文書中で上記辞書に存在しない未知語の有無を認識し、該未知語が存在しない場合には上記辞書のみを用いて翻訳を実行し、該未知語が存在した場合にはネットワークに接続された他の情報処理装置が有する辞書への検索要求を行い、その結果得られた情報を用いて上記翻訳対象文書の翻訳を完成させるとを特徴とする機械翻訳方法。

【請求項6】ある見出し語に関する辞書検索要求を受け取り、該見出し語に関する辞書情報をサーバ辞書を検索して取得し、該見出し語に関する辞書情報が前記サーバ辞書に存在しない場合には、他辞書サーバに検索要求を行い、他辞書サーバから取得した辞書情報を返送する分散辞書検索方法において、過去の検索要求に関する情報に基づいて他辞書サーバの優先順位を決定し、該優先順位に従って他辞書サーバに検索要求を行うことを特徴と

する分散辞書検索方法。

【請求項7】ある見出し語に関する辞書検索要求を受け取り、該見出し語に関する辞書情報をサーバ辞書を検索して取得し、該見出し語に関する辞書情報が前記サーバ辞書に存在しない場合には、他辞書サーバに検索要求を行い、他辞書サーバから取得した辞書情報を返送する分散辞書検索方法において、翻訳対象文書中に出現する単語に基づいて他辞書サーバの優先順位を決定し、該優先順位に従って他辞書サーバに検索要求を行うことを特徴とする分散辞書検索方法。

【請求項8】上記他辞書サーバの優先順位決定方法は、上記翻訳対象文書中に出現する単語の分野情報に基づいて上記翻訳対象文書の分野を推定し、該推定された分野における過去の検索要求に関する情報に基づいて他辞書サーバの優先順位を決定することを特徴とする請求項7に記載の分散辞書検索方法。

【請求項9】上記他辞書サーバの優先順位決定方法は、上記翻訳対象文書中に出現する少なくとも一部の単語の辞書情報を保持する他辞書サーバの情報を保持し、該他辞書サーバの情報および上記翻訳対象文書中の単語を用いて他辞書サーバの優先順位を決定することを特徴とする請求項7に記載の分散辞書検索方法。

【請求項10】ネットワークに接続された文書を他の自然言語に翻訳するネットワーク分散型機械翻訳システムにおいて、複数の辞書サーバを有し、各々の辞書サーバは、辞書とネットワークを介してアクセス可能な文書の統計情報に基づいて作成された分散辞書インデックスを有し、上記文書を含む語句の見出し語が上記辞書に存在しない場合には、上記分散辞書インデックスを参照して他の辞書サーバより該見出し語に関する情報を得て、該文書の翻訳を実行することを特徴とするネットワーク分

散型機械翻訳システム。

【請求項11】ネットワークに接続された文書を他の自然言語に翻訳するネットワーク分散型機械翻訳システムにおいて、複数の辞書サーバを有し、各々の辞書サーバは、辞書とネットワークを介してアクセス可能な文書のアクセス頻度情報に基づいて作成された分散辞書インデックスを有し、上記文書を含む語句の見出し語が上記辞書に存在しない場合には、上記分散辞書インデックスを参照して他の辞書サーバより該見出し語に関する情報を得て、該文書の翻訳を実行することを特徴とするネットワーク分散型機械翻訳システム。

【請求項12】辞書を参照しながら自然言語によって表される文書を他の種類の自然言語に翻訳するコンピュータシステムで用いられるコンピュータ可読媒体であって、自然言語で表された翻訳対象文書を読み込むための第1の命令コードと、上記翻訳対象文書から辞書引き対象語句を抽出するための第2の命令コードと、上記辞書引き対象語句の意味を予め準備された辞書から得て他の自然言語に翻訳するための第3の命令コードと、上記辞書が上記辞書引き対象語句の意味を保持しないときに、該辞書引き対象語句を未知語と認定して該未知語部分の翻訳を停止したまま後続部分の翻訳を継続させるための第4の命令コードと、ネットワークで接続された他の情報端末に対して上記未知語に関する情報を問い合わせるための第5の命令コードと、上記他の情報端末から得られた上記未知語に関する情報によって、上記未知語部分の翻訳を確定し、他の翻訳部分と併せて翻訳結果とするための第6の命令コードとを有するコンピュータプログラムを保持することを特徴とするコンピュータ可読媒体。